

# Scholarly Publications beyond Pay-walls. Increased Citation Advantage for Open Publishing

Susanne Mikki

## Abstract

### Purpose

First, we aim to determine the total amount of scholarly articles freely available on the internet.

Second, we aim to prove whether there exists a citation advantage for open publishing.

### Design

The total scholarly publication output of Norway is indexed in Ceres, the Current Information System in Norway. Based on these data, we searched Google Scholar by either DOIs or titles and denoted a document as open available (OAv), when a link to a full-text was provided. We analysed the extracted data by publishing year, citations, availability and provider. Based on additional information indexed by Ceres, we furthermore analysed the data by year, institution, publisher and discipline.

### Findings

We find that the total share of freely available articles is 68 %. Articles not available belong to prestigious publishers such as Elsevier, Springer, Routledge and Universitetsforlaget (the largest Norwegian academic publisher), which may be particularly essential for scholars worldwide. The largest provider, according to Google Scholar's main link provision, is ResearchGate. In addition, institutional repositories play a major role in posting free article versions. Articles belonging to natural sciences and technology, and medicine and health were more likely to be open than articles

belonging to the social sciences and humanities. Their respective OAv-shares are 72 %, 58 % and 55 %.

We find a clear citation advantage for open publishing; on average, these documents received twice as many citations, indicating that open access is the future in publishing.

### **Limitations**

This study is limited to scholarly articles only. Books and book chapters, which are usual publication formats for the humanities and social sciences, are excluded. Results do therefore not adequately reflect the situation for these disciplines.

Furthermore, this study is limited to documents freely available on the internet, independent of the “legal” status of the posted full-text. With the data at hand, we were not able to distinguish between gold, green, hybrid, purely pay-walled and illicitly posted documents.

### **Value**

Usually, articles indexed in Web of Science or SCOPUS are objects of investigation. However, these databases do not sufficiently cover the humanities and social sciences, and therefore cannot be representative of the total scholarly article output. This study captures the total article output of a country, independent on discipline and provides new insight into open publishing.

### **Introduction**

Open access is advocated widely within academia. It is declared through governmental policies and required by the main public funding agencies. The subject of this study is how the situation has evolved in Norway in particular.

In Norway, the Government's view is that publicly funded research must be made openly available (Ministry of Education and Research, 2013). In accordance with the governmental view, the Norwegian Research Council, the most important funder, requires that all scientific articles, wholly or partially funded by them, to be open by the day of publishing, either directly on the journal's website or indirectly through an institutional repository (Norwegian Research Council, 2014).

Universities and other research institutions are implementing local policies and practices to meet the new requirements. Until recently, the research communities have been reluctant to adopt open standards and have kept to traditional, pay-walled publishing. However, positive side effects, such as increased visibility and impact, and immediate feedback seem to change the game.

Based on our pilot study on climate (Mikki, Al Ruwehy, Gjesdal, & Zygmuntowska, accepted by *Library Hi Tech*), where we found 74% of all articles freely available, we wonder whether the free-share for the remaining subjects is equally high, and investigate here the total article free-share for Norway, independent of subject orientation.

The country's total scholarly publication output is registered in Ceres, formerly CRISTin, the Current Research Information System in Norway. Ceres' publication data is used for recall in Google Scholar (GS), a free online service for scholarly literature. Its benefit lies in its wide content coverage and smart full-text recognition.

The strength of this study lies in the completeness of the applied dataset. It covers all disciplines and is not restricted to content indexed by, e.g. Web of Science or SCOPUS, which are usually used for these types of studies. As long as an article is peer reviewed and scholarly approved (Universitets- og høyskolerådet, 2004), it is included, independent of language and subject orientation.

Open availability has been investigated for a while. Archambault et al. (2013) found above 40 % of all articles published between 2004 and 2011 were open. Similar results were obtained by a recent study regarding highly-cited documents (Martín-Martín, Orduna-Malea, Ayllón, & Delgado López-

Cózar, 2016). The case studies by Mikki et al. (accepted by *Library Hi Tech*), Jamali and Nabavi (2015) and Pitol and De Groot (2014) reported the highest shares so far, 74%, 61% and 70% respectively. These findings are limited to specific cases and sets of publications and call for further investigations and verification.

The question of open access citation advantage is as old as the possibility of publishing free online articles. Although the advantage is proved in much research (e.g. Antelman, 2004; Harnad & Brody, 2004; Swan, 2010), it is valuable to keep track of the new trends and to analyse different samples of open articles.

Whether there exists a citation advantage for open documents is of special interest to research communities. Such an advantage will benefit the authors and particularly challenge the traditional publishing industry. Recent studies (Archambault, Côté, Struck, & Voorons, 2016; Jamali & Nabavi, 2015; Mikki et al., accepted by *Library Hi Tech*) do indeed report a considerable (50%) higher citation impact for open documents. Since there does not exist a complete index on open available documents, these studies involve foremost case studies, which may not necessarily be universally representative (e.g. Hersh & Plume, 2016; Hua, Sun, Walsh, Worthington, & Glenny, 2016). They involve documents freely available on the internet, independent of their official open access status ("Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities," 2003). In order to add evidence to previous findings and to document an ongoing change in how scholars communicate, we aim at investigating the situation in Norway.

Based on the total article output in Norway, we want to

1. determine the total amount of scholarly articles freely available on the internet;
2. see whether there exists a citation advantage for open publishing.

## Methodology

Based on the total scholarly publication output of Norway (Current Research Information System in Norway), we searched Google Scholar for the documents' full-texts. Since open publishing foremost is limited to journal articles, only these are considered. We further limited the study to the five recent publishing years (2011–2015).

We define a document as open available (OAv) as long as it has been made freely accessible on the internet. This includes availability directly through the publisher's website, open repositories or academic services (e.g. ResearchGate, Academia, or institutional websites).

Our definition of open availability (abbreviated as OAv) must not be confused with the official definition of open access (usually abbreviated as OA). The official way of achieving open access is through publishing in open access journals (gold) and archiving in repositories (green); these ways insure long-term preservation and harvesting ("Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities," 2003). Open available copies provided by ResearchGate and similar sites are not covered by that definition, but by ours on open availability. The "legal" status of the document is irrelevant to this study, and data may include illicitly posted documents (i.e. posted at a time, location or in a version in conflict with the journal's policy).

We used either the unique Digital Object Identifier (DOI) or title for our Google Scholar searches and denoted a document as OAv, when a link to a full-text was listed. We did not verify whether the full-text was de facto available for each single item. Neither did we check whether the linked version is a pre-print version or the final publisher's version nor whether these two eventually differed. We performed the searches off campus to avoid access through our library SFX link resolvers to pay-walled items.

Automatic sampling was carried out by web scraping, extracting the following parameters:

Title, Authors, Publication Year, Cited by, format and full-text provider (Fig. 1). The extracted title was compared with the Ceres-title in order to exclude false matches.

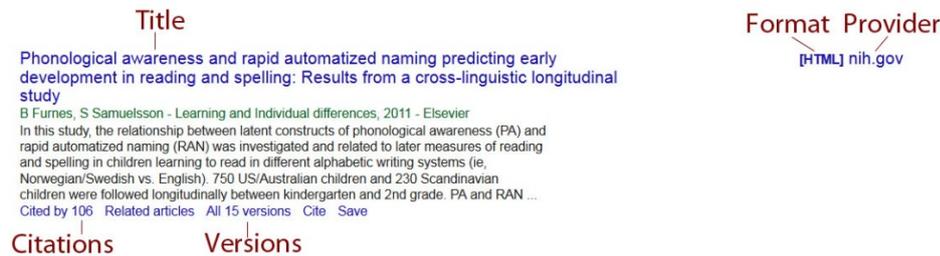


Fig. 1: Screen plot of Google Scholar’s search result; extracted fields are indicated

## Results

The total amount of scholarly articles in Norway between 2011 and 2015 comprises 70,882 items. Using Google Scholar, we retrieved 94 % of these (Table 1). Thus, only 6 % remained hidden. We found that these articles belong to journals with a foremost national orientation.

The majority of retrieved articles (54,064, 76 %) were recalled by their DOI. When recall by DOI failed or no DOI was available, the articles were recalled by their title. Titles were verified manually to assure that GS found the correct document and had not returned a false match. Articles which were assigned a DOI were more likely to be recalled (99 %) than articles without a DOI (74 %).

Table 1: Scholarly articles retrieved by Google Scholar (2011–2015).

Recalled articles	Number of articles	Percentage of Total
Yes	66691	94%
No	4191	6%
<b>Total</b>	<b>70882</b>	<b>100%</b>

The percentage of openly available articles is given in Fig. 2. The total OAv-share is 68 %. The share is highest for articles that were assigned a DOI (70 %). For articles searched by title, the percentage is 59 %, about ten percent lower.

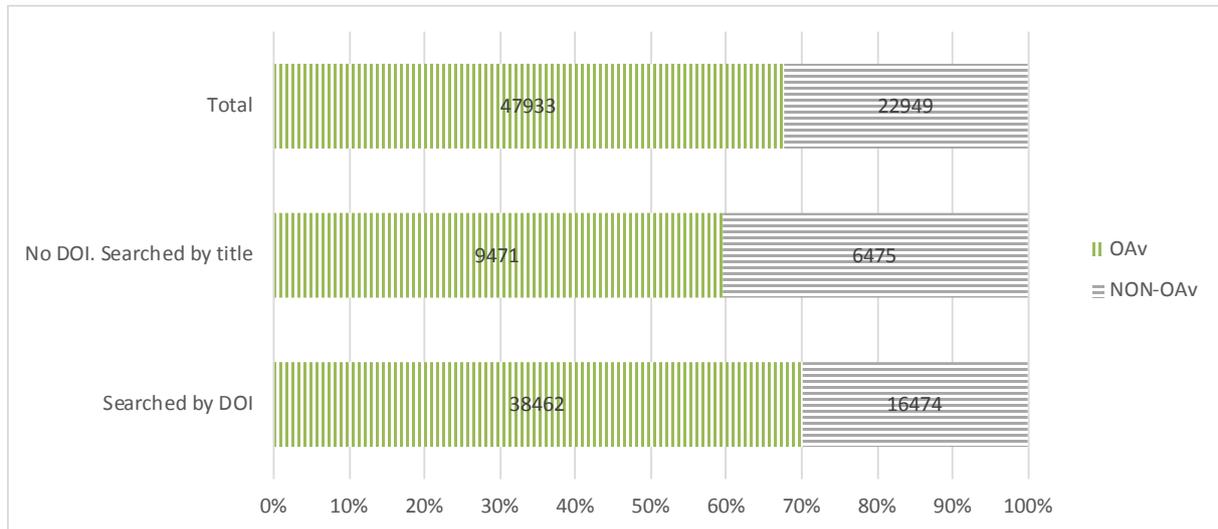


Fig. 2: Openly available articles searched in GS by either DOI or title (2011–2015).

For the investigated period, the share of open articles has slightly increased (Table 2). The decrease for the most recent year (2015, 68.3%) is most likely caused by publisher embargoes.

Table 2: Percentage of open available publications by year (2011–2015).

	OAv share
<b>2011</b>	66.9%
<b>2012</b>	67.9%
<b>2013</b>	69.1%
<b>2014</b>	71.4%
<b>2015</b>	68.3%

We determined the most frequent providers according to GS main full-text link, not considering possible parallel versions (compare Fig. 1). We did not check whether the provided articles indeed were fully available and in agreement with publisher’s policies on open access. Information was taken as given by GS. We found the academic network sites, such as ResearchGate and academia.edu

at the very top of the provider list (Table 3), which indicates that these services succeed both in making the researchers post their manuscripts and establishing platforms for sharing and discovering. Traditional publishers, such as Wiley, sciencedirect (Elsevier) and Springer, appear also among the most frequent providers. Articles belonging to these are either gold or hybrid. Providers with domain .no (10% in total) belong foremost to Norwegian open access repositories. We presume these articles to be in accordance to the publishers' open access policies.

**Table 3: Number of articles by provider, only the most frequent providers listed (2011–2015).**

<b>Provider</b>	<b>Number of articles</b>
<b>researchgate.net</b>	8644
<b>academia.edu</b>	2652
<b>bibsys.no</b>	2339
<b>biomedcentral.com</b>	2149
<b>wiley.com</b>	1987
<b>sciencedirect.com</b>	1784
<b>arxiv.org</b>	1671
<b>plos.org</b>	1394
<b>oxfordjournals.org</b>	1007
<b>nih.gov</b>	1006
<b>springer.com</b>	969
<b>uio.no</b>	877
<b>psu.edu</b>	711
<b>uit.no</b>	691
<b>diva-portal.org</b>	674
<b>tandfonline.com</b>	582
<b>acs.org</b>	578
<b>hio.no</b>	554
<b>hindawi.com</b>	523
<b>semantic scholar.org</b>	506
<b>bmj.com</b>	485
<b>uib.no</b>	453
<b>nature.com</b>	452
<b>ntnu.no</b>	392

A free article share of about 70% is impressive. However, the remaining 30% represent a crucial part of the scholarly output and belong to the most prestigious publishers. Based on Ceres' journal list (hosted by NSD), we were able to determine articles by publisher. Elsevier, Springer, Routledge and

Universitetsforlaget (Norwegian) are those with the highest number of pay-walled articles (Fig. 3). Routledge and Universitetsforlaget are publishers with the lowest OAv-shares. More than 70 % of their articles were pay-walled; corresponding numbers for Elsevier are 35 % and for Springer 32 %.

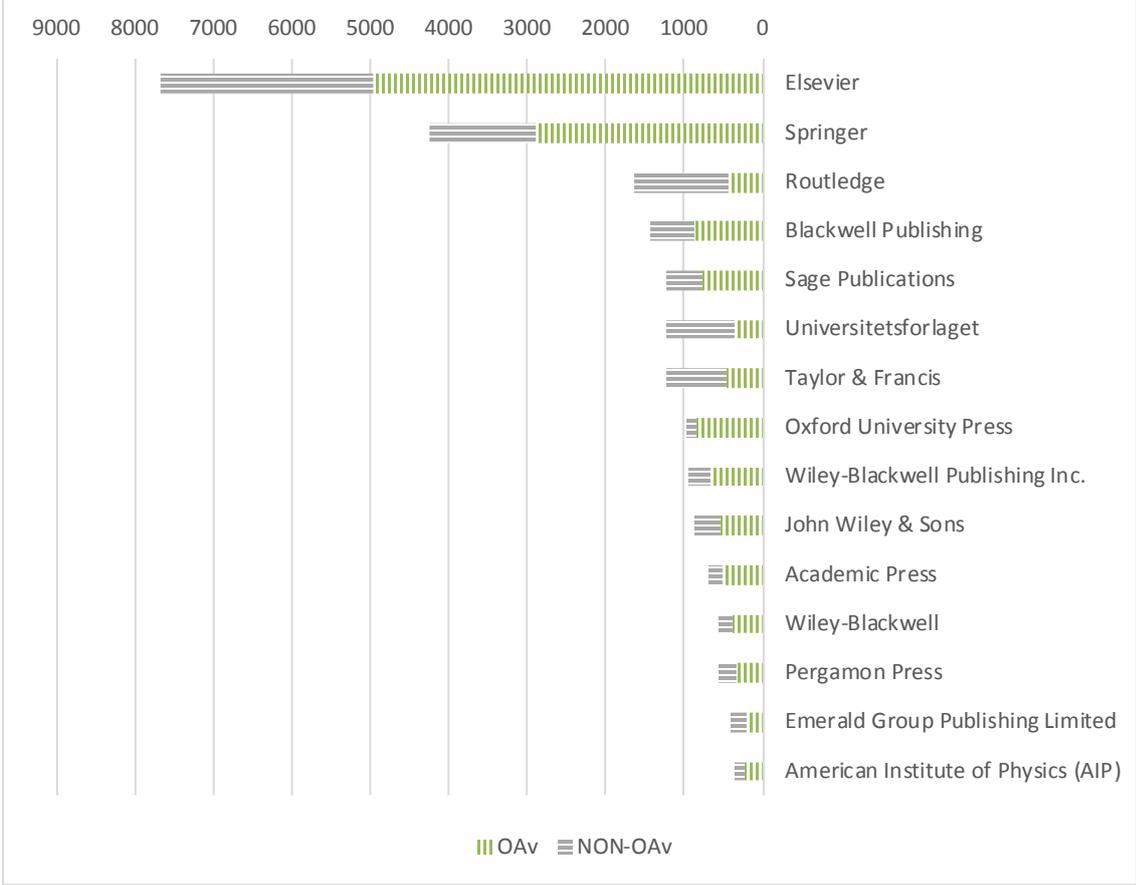


Fig. 3: Number of articles by publisher (top 15) and their status of availability.

The applied journal list (Ceres/NSD) also includes information on the main discipline of a journal. According to that information, the highest numbers of open articles belong to the natural sciences and technology, followed by medicine and health (Fig. 4). The OAv-shares for both of these disciplines are 72 %. The numbers for the social sciences and humanities are considerably lower (58 % and 55 % respectively). We find indeed pronounced differences between these main disciplines, and the tendency towards openness seems to depend on them. However, within disciplines, patterns may vary by sub-field.

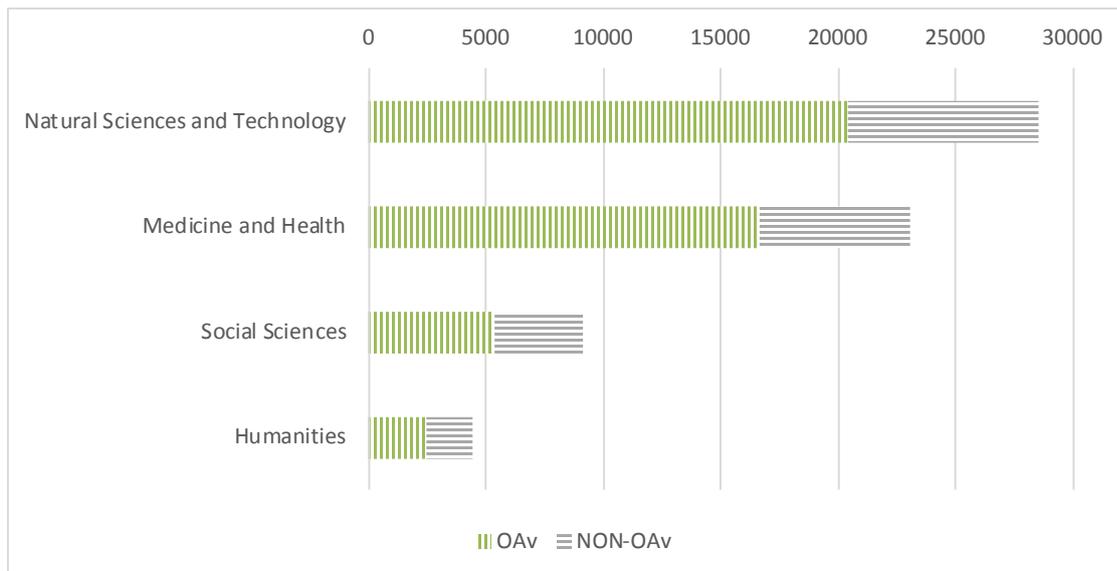


Fig. 4: Open available articles by discipline.

We also investigated the OAv-shares by institution and find that the largest universities hold the highest shares with the University of Bergen and The Arctic University of Norway (UiT) at the top, 73 % for both (Fig. 5). Whether this is an effect of allocated institutional funds for open access publishing and systematic institutional archiving, is unknown. The awareness towards open access publishing is high at the examined institutions, not least because of requirements by external funding bodies. The university with the lowest OAv-share, of the four shown in Fig. 5, is NTNU. Whether this is caused by the NTNU's subject profile, with weight on engineering and technology, or less awareness and dedication towards open access publishing, is not clear.

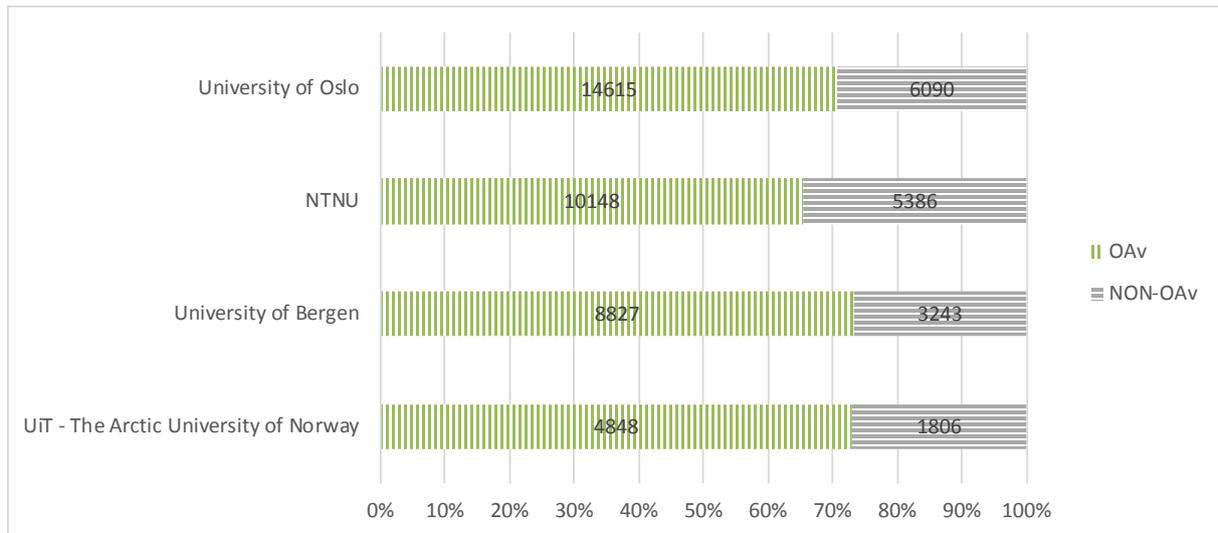


Fig. 5: Publications by availability (2011–2015) at the four largest universities of Norway.

Retrieved citations are accumulated from the day of publishing/posting until the day of searching (early January 2017) with an applied citation window of 1–6 years. We used the citation median, the citation average, and the h-index, as indicators of scholarly impact (Fig. 6). All the three indicators show similar patterns. We clearly see that the impact for both open and pay-walled documents increases with age of the article. For example, the citation median in 2015 to open articles was equal to 3 citations and increases linearly with 3 citations by year, resulting in 15 citations in 2011. For open articles, the impact is clearly higher compared to pay-walled articles. The citation median is more than twice (2.3), the citation average 1.8 and the h-index 2.3 the value of pay-walled items. Hence all these three citation measures indicate a marked higher impact of open articles.

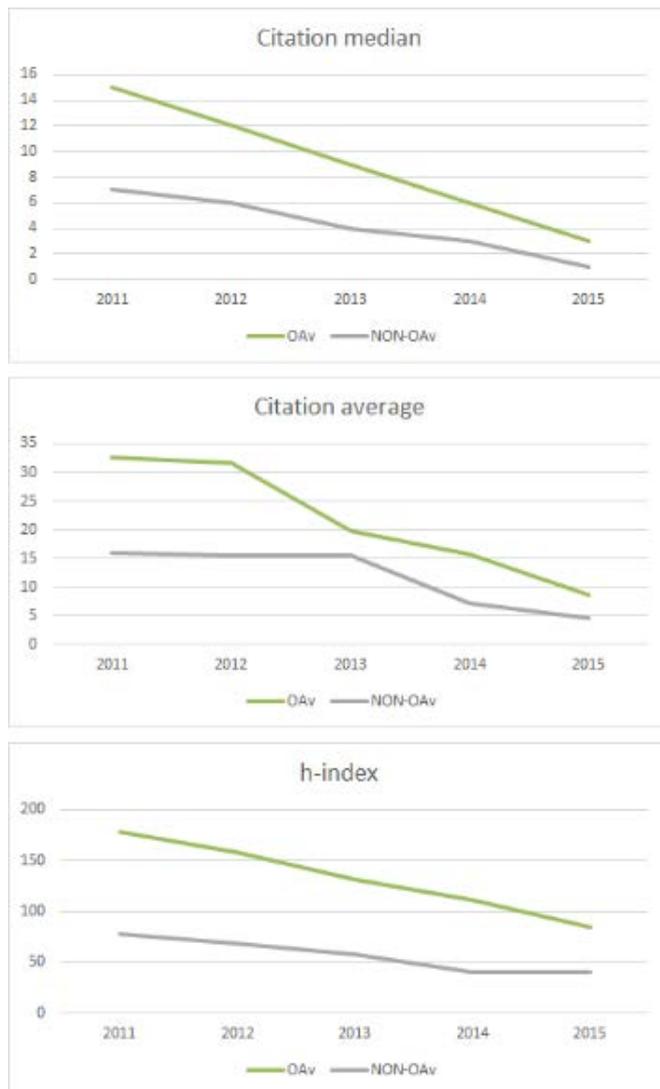


Fig. 6: Citation median, citation average and h-index by year for open (OAv) and pay-walled (NON-OAv) articles.

## Discussion and conclusions

Based on Norway's total scholarly article output, as indexed by Ceres, we found one of the highest so far reported open shares. Independent of discipline, 68% of all articles were openly available. The shares varied slightly by institution and the University of Bergen together with The Arctic University of Norway (UiT) were on top with 73%.

Open availability depends strongly on discipline, and the highest proportion was found for the natural sciences and technology and medicine and health sciences (72 %). These results are in accordance with our earlier findings (Mikki et al., accepted by *Library Hi Tech*) regarding climate impact on ancient societies (74%), and findings by Jamali and Nabavi (2015) who reported an OAv-share between 60% and 70%, dependent on sub-field. According to these studies, climate-related publications seem to be ahead when it comes to open publishing.

We find the OAv-shares for the social sciences (58 %) and for the humanities (55 %), for the social sciences slightly lower than reported by Jamali and Nabavi (2015) (60.8 %). The deviation can be explained by the different data sets. In this study, articles indexed by Ceres were applied, including national and non-English publishers, which are less likely to be captured by GS. The study of Jamali and Nabavi on the other side, was based only on items indexed by GS.

According to the GS link provider, off Campus, and disregarding possible alternative versions, ResearchGate and Academia were the services which most frequently offered a free version. Our findings underpin the importance of these user-driven services and their ability to turn themselves into academic market and network places. Norwegian sites (domain .no) represent mainly institutional repositories and comprise about 10% of the provided full-texts. This share is relatively low and hardly reflects the total number of archived documents. Additional archived copies most likely exist. However, as far as a journal publisher version is freely available, GS links most likely to this original version (see also Pitol & De Groot, 2014). The degree to which GS provides links to academic network sites above institutional repositories is worth investigating more closely. We find that archiving plays a crucial role in providing full-texts, and requirements by strong funding agencies are main drivers in this regard (European Commission, 2013; NFR, 2014).

For the investigated period (2011–2015), we observed a slight increase of open available documents for all years except the last (2015), when it dropped. Obviously, publisher-imposed embargoes, but also reluctance of the authors and time-demanding administration procedures, delay the point for

when the document is posted on the internet (confer Archambault et al., 2016). Embargoes are meant to protect the publishers' financial interests and usually last between 6 and 24 months after initial publication (Sutton, 2013). However, the traditional publishing model, with closed access and overpriced subscription rates, seems to be no longer sustainable (Chen, 2016), as the citation advantage for open documents clearly indicates.

We used the citation median, the citation average and the h-index as indicators of scholarly impact. All the indicators show similar patterns and a pronounced higher impact for open articles. Both the citation median and the citation average are more than twice (2.3) the value of pay-walled articles. These findings are in accordance with findings by the previously-mentioned studies (Archambault et al., 2016; Jamali & Nabavi, 2015; Pitol & De Groote, 2014) and confirm in fact an increased citation advantage of open documents. We conclude that open publishing is a good strategy for maximizing research impact.

A free-article share of 68 % as reported here may sound impressive. However, the remaining pay-walled articles belong to the most prestigious publishers such as Elsevier, Springer, Routledge and Universitetsforlaget (the largest Norwegian academic publisher) with relatively restrictive open access policies.

## Acknowledgements

I want to thank my colleagues Hemed Al Ruwehy and Øyvind Gjersdal who kindly provided the data for this study.

## References

- Antelman, K. (2004). Do open-access articles have a greater research impact? *College & Research Libraries*, 65(5), 372. Retrieved from <http://crl.acrl.org/index.php/crl/article/view/15683>
- Archambault, E., Amyot, D., Deschamps, P., Nicol, A., Rebout, L., & Roberge, G. (2013). *Proportion of Open Access Peer-Reviewed Papers at the European and World Levels—2004-2011* Vol. 1. *Science-Metrix* (pp. 495.6505). Retrieved from [http://www.science-metrix.com/pdf/SM\\_EC\\_OA\\_Availability\\_2004-2011.pdf](http://www.science-metrix.com/pdf/SM_EC_OA_Availability_2004-2011.pdf)

- Archambault, E., Côté, G., Struck, B., & Voorons, M. (2016). Research impact of paywalled versus open access papers. *Science-Metrix and 1science*. Retrieved from <http://www.1science.com/oanumbr.html>
- Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities. (2003). Retrieved from <https://openaccess.mpg.de/Berlin-Declaration>
- Chen, X. (2016). A Middle-of-the-Road Proposal amid the Sci-Hub Controversy: Share “Unofficial” Copies of Articles without Embargo, Legally. *Publications*, 4(4), 29.
- European Commission. (2013). *Horizon 2020, Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020*. Retrieved from [https://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-pilot-guide\\_en.pdf](https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf).
- Harnad, S., & Brody, T. (2004). Comparing the impact of open access (OA) vs. non-OA articles in the same journals. *D-lib Magazine*, 10(6), 9.
- Hersh, G., & Plume, A. (2016). Citation metrics and open access: what do we know? Retrieved from <https://www.elsevier.com/connect/citation-metrics-and-open-access-what-do-we-know>
- Hua, F., Sun, H. Y., Walsh, T., Worthington, H., & Glenny, A. M. (2016). Open access to journal articles in dentistry: Prevalence and citation impact. *Journal of Dentistry*, 47, 41-48. doi:10.1016/j.jdent.2016.02.005
- Jamali, H. R., & Nabavi, M. (2015). Open access and sources of full-text articles in Google Scholar in different subject fields. *Scientometrics*, 105(3), 1635-1651. doi:10.1007/s11192-015-1642-2
- Martín-Martín, A., Orduna-Malea, E., Ayllón, J. M., & Delgado López-Cózar, E. (2016). A two-sided academic landscape: portrait of highly-cited documents in Google Scholar (1950-2013). *Revista Española de Documentación Científica, Preprint*. doi:10.3989/redc.2016.4.1405
- Mikki, S., Al Ruwehy, H. A., Gjesdal, Ø. L., & Zygmuntowska, M. (accepted by Library Hi Tech). Filter bubbles in interdisciplinary research. A Case study on climate and society.
- Ministry of Education and Research. (2013). *Meld. St. 18 (2012-2013). Long-term perspectives – knowledge provides opportunity Norway*: Published under: Stoltenberg's 2nd Government Retrieved from <https://www.regjeringen.no/en/dokumenter/meld.-st.-18-2012-2013/id716040/>.
- NFR. (2014). *Tilgjengeliggjøring av forskningsdata. Policy for Norges forskningsråd* Retrieved from <http://www.forskningsradet.no/servlet/Satellite/?SSURlapptype=BlobServer&blobkey=id&SSURlcontainer=Default&blobwhere=1274506370325&SSURlsession=false&blobheader=application%2Fpdf&ssbinary=true&blobheadername1=Content-Disposition%3A&blobheadervalue1=+attachment%3B+filename%3D9788212033610.pdf&SSURlsscontext=Satellite+Server&blobcol=urldata&blobtable=MungoBlobs#satellitefragment>
- Norwegian Research Council. (2014). *The Research Council's Principles for Open Access to Scientific Publications*. Retrieved from <http://www.forskningsradet.no/servlet/Satellite?blobcol=urldata&blobheader=application%2Fpdf&blobheadername1=Content-Disposition&blobheadervalue1=+attachment%3B+filename%3D%22RCNpolicyOpenAccess.pdf%22&blobkey=id&blobtable=MungoBlobs&blobwhere=1274506495121&ssbinary=true>.
- NSD. Norwegian Register for Scientific Journals, Series and Publishers. Retrieved from [https://dbh.nsd.uib.no/publiseringsskanaler/Forside.action?request\\_locale=en](https://dbh.nsd.uib.no/publiseringsskanaler/Forside.action?request_locale=en)
- Pitol, S. P., & De Groote, S. L. (2014). Google Scholar versions: do more versions of an article mean greater impact? *Library Hi Tech*, 32(4), 594-611. doi:10.1108/LHT-05-2014-0039
- Sutton, S. C. (2013). Open access, publisher embargoes, and the voluntary nature of scholarship An analysis. *College & Research Libraries News*, 74(9), 468-472.
- Swan, A. (2010). The Open Access citation advantage: Studies and results to date. Retrieved from <https://eprints.soton.ac.uk/268516/>
- Universitets- og høyskolerådet. (2004). *Vekt på forskning: nytt system for dokumentasjon av vitenskapelig publisering : innstilling fra faglig og teknisk utvalg til UHR* (pp. 83 s.). Retrieved from [http://www.uhr.no/documents/Vekt\\_p\\_forskning\\_sluttrapport.pdf](http://www.uhr.no/documents/Vekt_p_forskning_sluttrapport.pdf)

